Deep Learning Classification Models Built with Two-step Transfer Learning for Age Related Macular Degeneration Diagnosis

Guangzhou An, Hideo Yokota, Naohiro Motozawa, Seiji Takagi, Michiko Mandai, Shohei Kitahata, Yasukiko Hirami, Masayo Takahashi, Yasuo Kurimoto, and Masahiro Akiba

Abstract— **The objective of this study was to build deep learning models with optical coherence tomography (OCT) images to classify normal and age related macular degeneration (AMD), AMD with fluid, and AMD without any fluid. In this study, 185 normal OCT images from 49 normal subjects, 535 OCT images of AMD with fluid, and 514 OCT mages of AMD without fluid from 120 AMD eyes as training data, while 49 normal images from 25 normal eyes, 188 AMD OCT images with fluid and 154 AMD images without any fluid from 77 AMD eyes as test data, were enrolled. Data augmentation was applied to increase the number of images to build deep learning models. Totally, two classification models were built in two steps. In the first step, a VGG16 model pre-trained on ImageNet dataset was transfer learned to classify normal and AMD, including AMD with fluid and/or without any fluid. Then, in the second step, the fine-tuned model in the first step was transfer learned again to distinguish the images of AMD with fluid from the ones without any fluid. With the first model, normal and AMD OCT images were classified with 0.999 area under receiver operating characteristic curve (AUC), and 99.2% accuracy. With the second model, AMD with the presence of any fluid, and AMD without fluid were classified with 0.992 AUC, and 95.1% accuracy. Compared with a transfer learned VGG16 model pre-trained on ImageNet dataset, to classify the three categories directly, higher classification performance was achieved with our notable approach. Conclusively, two classification models for AMD clinical practice were built with high classification performance, and these models should help improve the early diagnosis and treatment for AMD.**

I. INTRODUCTION

Age related macular degeneration (AMD) is the leading cause of visual loss among older people and the patient number is growing rapidly with aging society in developed countries [1]. AMD could be categorized into dry and wet AMD based on the absence of neovascularization or not [2]. In wet AMD, fluid leakage or bleeding from permeable capillaries network in sub-RPE, and following fibrotic scar in sub-retinal areas, derive severe photoreceptor degeneration [3]. It is very important to monitor AMD patients regularly with anti-VEGF therapies to retain a good visual acuity.

Optical coherence tomography (OCT) is commonly used ophthalmologic instrument which clearly describe particular pathology of AMD such as drusen, intra-retinal fluid (IRF) sub-retinal fluid (SRF), sub-retinal hyper-reflective material and retinal pigment epithelium detachment [4]. Among them, absence of IRF and/or SRF is the very important interpretation point for most doctors as the therapeutic initiation of anti-VEGF therapy and evaluation of its effect [5]. However, these increased requirements of interpretation on huge amount of OCT data are a big burden for doctors [6]. It is required by the doctors to develop automatic analyzing methods to screen and provide an indication for applying and evaluating the anti-VEGF therapy effect, with classifying normal, AMD with fluid, and AMD without fluid.

Recently, machine learning technology especially deep learning has seen dramatic progress, and has enabled the development of new algorithms for automate diagnosis of eye disease including AMD, glaucoma, diabetic retinopathy with OCT images or fundus photography as input [7,8,9,10,11,12,13]. In previous papers, there was no classification model reported for classifying normal, AMD with fluid, and AMD without fluid. It is difficult to judge absence of fluid with not obvious differentiation between them.

In this study, we presented a notable method of building deep learning classification models with OCT images to classify normal and AMD, and distinguish image of AMD with fluid from AMD without any fluid, together.

II. METHODS

A. Subjects and preprocessing of OCT images

This study enrolled 120 eyes of 120 AMD patients, and 49 eyes from 49 normal subjects, as training data group, and enrolled another group as test data group, including 77 eyes of 77 AMD patients, and 25 eyes from 25 normal subjects. The protocols of this study were approved by the institutional review board of RIKEN (Wako3 26-4).

OCT images of subjects in both group were captured with Heidelberg Spectralis OCT device, which is spectral domain OCT, in protocol of either radial-scan with 6.0 mm scan length or cross-scan with 9.0 mm scan length. In cross scan images, the central 6.0 mm area were cropped to have the same scan length as radial-scans and resized them into 496*496 pixels. As a result, there were185 normal OCT images, 535 OCT images of AMD with fluid, and 514 OCT mages of AMD

*This research is a joint research project of Topcon Corporation, RIKEN, and Kobe City Eye Hospital.

G.A. is with Topcon Corporation, Tokyo, JAPAN, and RIKEN Center for Advanced Photonics, Wako, JAPAN, and Kobe University Graduate School of System Informatics, Kobe, JAPAN.

M.A. is with Topcon Corporation, Tokyo, JAPAN, and RIKEN Center for Advanced Photonics, Wako, JAPAN

H.Y. is with RIKEN Center for Advanced Photonics, JAPAN, and Kobe University Graduate School of System Informatics, Kobe, JAPAN.

N.M. is with Kobe City Eye Hospital, Kobe, JAPAN, RIKEN Center for Biosystems Dynamics Research, Kobe, JAPAN, and Kyoto University Graduate School of Medicine, Kyoto, JAPAN.

S.T. is with Kobe City Eye Hospital, Kobe, JAPAN, and Teikyo University Hospital Mizonokuchi, Kawasaki, JAPAN.

M.M., S.K., Y.H., M.T., and Y.K. are with Kobe City Eye Hospital, Kobe, JAPAN, and RIKEN Center for Biosystems Dynamics Research, Kobe, JAPAN.

without fluid in training data, while in test data, there were 49 normal images, 188 AMD OCT images with fluid and 154 AMD images without any fluid. To increase the number of training data, 3 images from each OCT image were cropped from left, middle, right side, with the size of 224*224pixels, while the vertical position of cropping center is on the RPE line, which is detected automatically. All these cropped images were reviewed by three ophthalmologists independently, labelled as normal, AMD with fluid, and AMD without fluid (Fig. 1). Only the images with the same results by the graders were selected for training or validating the classification models. Finally, as training data, 476 normal images, 1,145 images of AMD with fluid and 1,026 mages of AMD without fluid were included, while in test data, 134 normal images, 402 AMD OCT images with fluid and 347 AMD without any fluid were included.

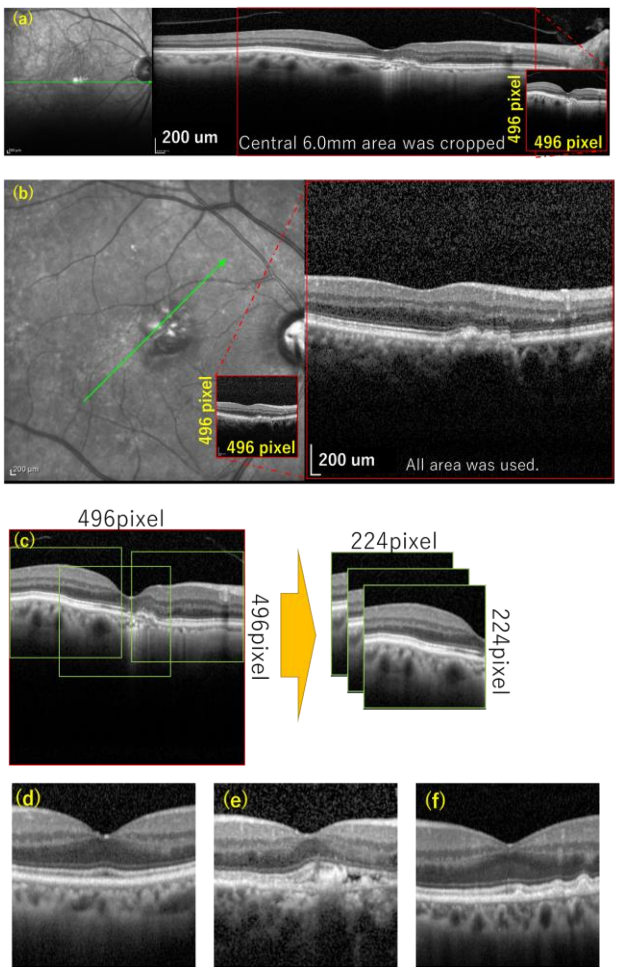

Figure 1. Preparation for deep learning. (a) Preprocessing for cross scan OCT images. (b) Preprocessing for radial cross scan OCT images. (c) Cropping from preprocessed OCT images, (d) A sample of cropped OCT image labelled as normal. (e) A sample of cropped OCT image labelled as AMD with fluid. (f) A sample of cropped OCT image labelled as AMD without fluid.

B. *CNN Architecture and Transfer Learning*

A convolutional neural network (CNN) is a supervised classifier based on deep learning [8]. It is comprised of several convolutional layers as well as subsampling layers, good at designing powerful filters to get the sensitive image features for the classification task. Until now, there are hundreds of CNN architectures, which got the top-performance for the ImageNet challenge. In this study, we adopted a CNN architecture VGG16, a CNN model with 16 layers being widely used to solve image classification task, to classify normal and AMD OCT images or to classify AMD with fluid, and AMD without fluid [9].

On the other hand, transfer learning is a machine learning method to apply a developed model for previous tasks to a new task domain. Based on this strategy, we transfer learned a VGG16 models to solve the classification task. Dropout, and data augmentation including horizontal flip, random rotation, and random shift, were used during training. An epoch number of 100 was used during training phase, and finally, from 100 VGG16 models, we select the model with the highest AUC.

C. *Proposed Approach*

Ophthalmologists always do the clinical judgement relevant with treatment after diagnosing whether it is disease or not. Similarly, in our proposed approach, we first transfer learned a VGG16 model pre-trained on ImageNet dataset, to classify normal and AMD images with training data, considering AMD with fluid and AMD without fluid as AMD category. Then, this transfer learned model was further transfer learned with OCT images of AMD with fluid, AMD without fluid in training data, to build the second model for judging whether there is fluid (Fig. 2).

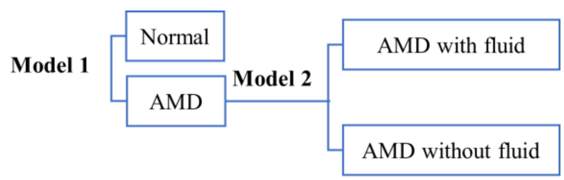

Figure 2. Proposed approach.

D. *Evaluation of Classification Performance*

A model for classifying the 3 categories (normal, AMD with fluid, and AMD without fluid) directly was built to compare with our approach. In this case, a VGG16 model pretrained on ImageNet dataset was also applied, and that model was transfer learned with the training data (Fig. 3).

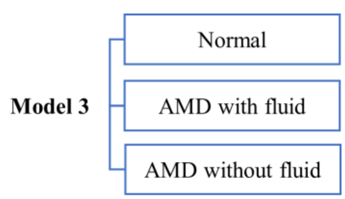

Figure 3. Comparison approach.

As an evaluation index of classification, the area under receiver operating characteristic curve (AUC) for the ability of VGG16 models was used. Besides AUC, we also examined accuracy of judgment for the original image, we combined judgment of each cropped image. In the first model, if there was an image of AMD of more than one, it was judged as AMD, otherwise it was judged as normal. The judgment of the AMD with fluid is done similarly, if there was more than one image of fluid of 3, it was judged as fluid image, or it was AMD without fluid. To compare the classification performance of classifying AMD and normal, in the third model, considering AMD with fluid and AMD without fluid as AMD category, the test data was input into the model, and calculated on the AUC and accuracy. Furthermore, to compare the ability of classifying AMD with fluid, and AMD without fluid OCT images were input into the third model to calculate the AUC and accuracy.

III. RESULTS

In this report, deep learning models with OCT images to classify normal and AMD, and distinguish AMD with fluid from AMD without any fluid. For classifying normal versus AMD images, AUC was 0.999, and accuracy was 99.2%. For the classification task of AMD with fluid and AMD without fluid, the classification performance were 0.992 AUC, and 95.1% accuracy. Compared with the model created by the comparison approach, in both cases, our proposed approach achieved higher performance (TABLE I).

TABLE I. CLASSIFICATION PERFORMANCE OF BUILT MODELS

Classification	**Proposed approach**		**Comparison approach**	
	AUC	***Accuracy***	***AUC***	***Accuracy***
Normal versus AMD (with fluid and without fluid)	0.999	99.2%	0.994	98.9%
AMD with fluid versus AMD without fluid	0.992	95.1%	0.988	94.2%

IV. DISCUSSIONS

In this study, we found that a transfer learning of VGG16 was suitable deep learning technique for automate screening of AMD, and judging AMD with fluid or not.

In the first model, the performance of classification for normal and AMD OCT images were 0.999 AUC, and 99.2% accuracy. In recent studies' report, deep learning method could achieve high accuracy in screening AMD from normal with OCT images. With OCT images as input, a CNN model pre-trained with ImageNet dataset, was transfer learned with 1,012 B-Scan OCT images was able to distinguish normal from AMD images with 96% accuracy. Our result was consistent with recent reports [14]. Data shuffling was applied twice on sum of training and test data group, to create two different sets of training and test data, to evaluate performance stability of our models created by our proposed approach. In the additional experiments, the AUCs of the models to classify normal versus AMD (with fluid and without fluid) using our proposed approach, were both over 0.999, and were better than the AUCs of the models created with the comparison approach.

In the second model, we could distinguish AMD with fluid from AMD without any fluid, with 0.992 AUC, and 95.1% accuracy. Detecting the presence of fluid at macular is the most important point for many ophthalmologists in therapeutic indication. Treatment of anti-VEGF injection therapy and redosing criteria depend on the whether there is IRF and/or SRF or not. Recent reports have shown that the amount of fluid can be quantified accurately and clearly recognizing the differentiation using deep learning method, and classify indicators of fluid in OCT images, which are key points for initial and anti-VEGF therapy decisions in AMD with 92% sensitivity, 91% specificity and 93% accuracy. Our result was consistent with this report. In our second model, the transfer learning was applied. In transfer learning, we could obtain high accuracy result with fewer dataset. It is considered to be effective when image data is limited, especially in clinic data. However, in most cases, deep learning requires huge amount of data. Recently, effectiveness of building deep learning models with transfer learning from pretrained ones has been reported [14]. Similar with the first model, data shuffling was applied twice on sum of training and test data group, to create two different sets of training and test data, to evaluate performance stability of our models created by our proposed approach. In the additional experiments, the AUCs of the models to classify AMD with fluid and AMD without fluid using the proposed approach, were both higher than 0.980, and were better than the AUCs of the models created with the comparison approach.

We applied grad-CAM, a commonly used method to find our models observe which area is important for VGG16 model to judge [15]. In detail, grad-CAM makes the area characteristic as a heat map. Therefore, it is possible for the doctors to confirm the important area of CNN classification models on each image. In detail, a heat-map for a classic AMD category indicating the effective region for the model to identify AMD was generated, while a heat-map was created with the discriminative region used by the second model to identifying wet AMD. Against a same OCT image, compared with heat-map created by the first model, the one created by the second model, the important area is just on the fluid, which is a quite important point to classify whether there is fluid or not in an OCT image (Fig. 4).

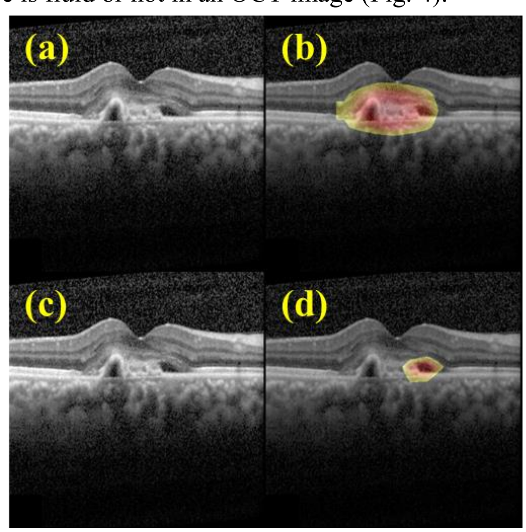

Figure 4. Important areas in our VGG16 models. A classic AMD with SRF was selected to confirm the heatmap for our VGG16 modelst. (a) Cropped AMD OCT image (b) Heatmap created by the first model to classify AMD and normal OCT images. (c) Cropped AMD OCT image, same as (a). (d) Heatmap created by the second model to classify AMD with fluid and AMD without fluid OCT images. Red regions correspond to high score for the classification.

Furthermore, we compared the heatmaps by model 2 (proposed approach) and model 3 (comparison approach) to identify AMD with fluid or not. The heatmaps by model 2 is more sensitive to find the effective area for the true classification. In the cropped OCT image of sample-1, model 2 was succeed to narrow the important area to the actual fluid area in OCT image, while in the cropped OCT image of sample-2, model 2 expanded the important area to the fluid area in OCT image, which might help to get a higher classification performance to classify the AMD with fluid and the AMD without fluid.

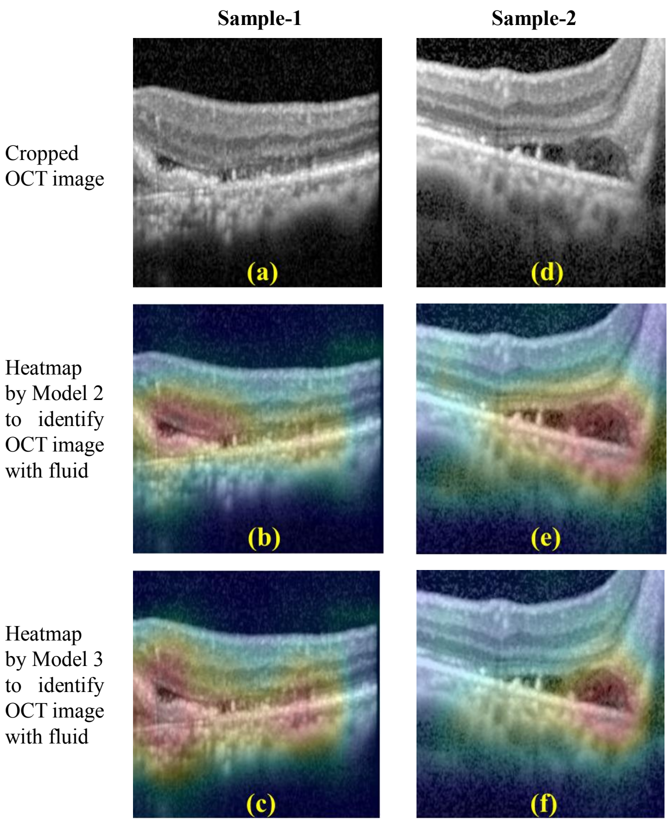

Figure 5. Differenciation of heatmaps created by built VGG16 models to identify AMD OCT image with fluid. Two cropped OCT images were randomly seleted to show the diffrenciation of heatmaps created by our VGG models to classify AMD with fluid and AMD without fluid, compared with the built model by the comparison approach. Red regions correspond to high score for the classification. (a) Cropped AMD OCT image with fluid from sample-1, (b) Heatmap created by Model 2 (propsed approach) for (a) to identify AMD with fluid, (c) Heatmap created by Model 3 (comparison approach) for (a) to identify AMD with fluid, (d) Cropped AMD OCT image with fluid from sample-2, (e) Heatmap created by Model 2 (propsed approach) for (d) to identify AMD with fluid, (f) Heatmap created by Model 3 (comparison approach) for (d) to identify AMD with fluid.

There are several limitations in this study. Only the images have been adopted that meet our research inclusion criteria. The images of low quality images and the patients with besides AMD were excluded. Grade of images which the three ophthalmologists did not agree on the presence of fluid were excluded. Therefore, it is unclear that, whether a VGG16 model, which was transfer learned with only these images, can be applied in other clinical applications. Furthermore, the resource of these OCT images is just a single academic center. In the future work, the research should be further done by including images from various OCT images from much more facilities, with multi-purpose.

V. Conclusion

Conclusively, two classification models for AMD clinical practice were built with high classification performance; in detail, one for classification of normal and AMD, and the other for classification of AMD with fluid and AMD without fluid, were built. These deep learning classification models should help improve the diagnosis of AMD, lead to better clinical AMD care. Furthermore, our findings of building deep learning classification model with transfer learning might help improve the classification performance in other application.

References

[1] L.S. Lim et al., “Age-related macular degeneration,” Ophthalmology, vol. 379, no. 9827, May 2012, pp. 1728–1738.

[2] A.C. Bird et al., “An international classification and grading system for age-related maculopathy and age-related macular degeneration. The International ARM Epidemiological Study Group,” Surv Ophthalmol., vol. 39, no. 5, 1995, pp. 367-374.

[3] I. Bhutto, and G. Lutty, “Understanding age-related macular degeneration (AMD): relationships between the photoreceptor/retinal pigment epithelium/Bruch's membrane/choriocapillaris complex,” Mol Aspects Med., vol. 33, no. 4, August 2012, pp. 295-317.

[4] C.V. Regatieri, L. Branchini, and J.S. Ducker, “The role of spectral-domain OCT in the diagnosis and management of neovascular age-related macular degeneration,” Ophthalmic Surg Lasers Imaging., vol. 42, Jul. 2012, Suppl: S56-S56.

[5] A.E. Fung et al., “An optical coherence tomography-guided, variable dosing regimen with intravitreal ranibizumab (Lucentis) for neovascular age-related macular degeneration,” Am J Ophthalmol. Apr., vol. 143, no. 4, 2007, pp. 566-583.

[6] M.R. Alexandru, and N.M. Alexandra, “Wet age related macular degeneration management and follow-up,” Rom J Ophthalmol., vol. 60, no. 1, 2016, pp. 9-13.

[7] U. Schmidt-Erfurth, et al., “Artificial intelligence in retina,” Progress in Retinal and Eye Research, vol. 67, 2018, pp. 1-29.

[8] S. Becker, and Y. “Lecun, “Improving the convergence of back-propagation learning with second-order methods,” Proceedings of the 1988 Connectionist Models Summer School, 1989, pp. 29-37.

[9] K. Simonyan, and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” Computer Vision and Pattern Recognition, 2015.

[10] M Caixinha, and S Nunes, “Machine Learning Techniques in Clinical Vision Sciences,” Curr Eye Res, vol. 42, 2017, pp. 1-15.

[11] P.M. Burlina et al., “Automated Grading of Age-Related Macular Degeneration From Color Fundus Images Using Deep Convolutional Neural Networks,” JAMA Ophthalmol., vol. 135, no. 11, 2017, pp. 1170-1176.

[12] C.S. Lee et al., “Deep Learning Is Effective for Classifying Normal versus Age-Related Macular Degeneration OCT Images,” Ophthalmol. Retina, vol. 1, no. 4, 2017, pp. 322-327.

[13] D.F. Jeffrey et al., “Clinically applicable deep learning for diagnosis and referral in retinal disease,” Nature Medicine, vol. 24, no. 9, 2018.

[14] M. Treder et al., “Automated detection of exudative age-related macular degeneration in spectral domain optical coherence tomography using deep learning,” Graefes Arch Clin Exp Ophthalmol., vol. 256, no. 2, 2018, pp. 259-265.

[15] R.S. Ramprasaath et al., “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization,” Computer Vision and Pattern Recognition, 2017.